DERIVE A LINEAR TRENDLINE

Author: Donald C. Braun
E-mail address: DonCBraun@aol.com

File: LinearTrendline.doc
Web address: https://explorersposts.grc.nasa.gov/post631/2011-2012/
Revision 1.6
February 3, 2012

(This document was originally created using Microsoft Word 2007/12 in Compatibility Mode, and Word 2010/14 in Compatibility Mode with Microsoft Equation Editor 3.0 .)

Section 1. Solve two linear equations for two unknowns by Cramer's rule

Reason or comment	Statement
	100) This document derives the least squares algorithm described in statement (**410**) to calculate a linear trendline passing near any number of given sample points.
Purpose of Section 1.	This Section 1 derives a method called **Cramer's rule** in theorem (**104**), which solves **two simultaneous linear equations in two variables** for the unique values of those two variables that satisfy both equations.
Define mathematical symbols & abbreviations.	**101**) The following table lists a few mathematical symbols and abbreviations used in this document.

Symbol or abbreviation	Meaning
$\equiv$	"is defined as" or "equals by substituting one or more definitions"
$\in$	"is an element of (the following set)" or "in (the following set)"
… or ⋯	"et cetera" or "etc." or "and so on" (The symbol itself is called an ellipsis.)
{ **…** }	a set whose elements are enclosed by the pair of braces ("curly brackets") (For example, $k \in \{\ 1, 3, 5, 7, 9\ \}$ means "k is one of five single-digit odd integers.")
def; defs	"definition" or "is defined as"; "definitions"
eqn; eqns	"equation"; "equations"
stmt; stmts	"statement"; "statements"
thm; thms	"theorem"; "theorems"

Reason or comment	Statement
Define a general **2×2 matrix** and present a specific numerical example.	**102**) A **2-by-2 matrix** is an array of 4 numerical values arranged in 2 horizontal rows by 2 vertical columns and enclosed by a pair of square brackets, such as $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ for any values A, B, C, and D; or for example $\begin{bmatrix} 3 & 5 \\ 7 & 9 \end{bmatrix}$.
Define the **determinant** of a 2×2 matrix.	**103**) The **determinant** of a 2×2 matrix is a numerical value characterizing the matrix and is typically denoted by enclosing the array of elements by vertical bars instead of by square brackets. The value of the determinants of the two matrix examples shown in the previous statement (**102**) are defined as
By definition of a determinant.	$\begin{vmatrix} A & B \\ C & D \end{vmatrix} \equiv A\,D - B\,C\,,$ and so specifically
An example where $A = 3$, $B = 5$, $C = 7$, and $D = 9$.	$\begin{vmatrix} 3 & 5 \\ 7 & 9 \end{vmatrix} \equiv (3)\,(9) - (5)\,(7) = 27 - 35 = -8\,.$

Reason or comment		Statement
An alternate notation for a 2×2 determinant is $\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} \equiv AD - BC$.		There is an extensive mathematical field of study called "**matrix theory**" or "**linear algebra**" that describes the properties and applications of matrices with arbitrary sizes and determinants of square matrices (that is, matrices whose number of rows equals the number of columns). In this document only the definitions (**102**) and (**103**) will be used in statements (**104**) and (**410**) below.
This theorem expresses **Cramer's rule** for solving 2 linear equations in 2 unknowns.	**104**)	By **Cramer's rule** (named after the Swiss mathematician Gabriel Cramer, who lived 1704-1752 but was not the first to state the rule), the values of two real variables x and y that satisfy both of the two simultaneous linear equations $Ax + By = C$ and $Fx + Gy = H$ can be calculated in terms of the known real constants A, B, C, F, G, and H by
Define D to be used for calculating x and y.		$D \equiv AG - BF \equiv \begin{vmatrix} A & B \\ F & G \end{vmatrix} \equiv$ **Cramer's denominator**,
These solutions for x and y satisfy both linear equations given above.		$x = \dfrac{CG - BH}{D} \equiv \dfrac{1}{D}\begin{vmatrix} C & B \\ H & G \end{vmatrix}$, and $y = \dfrac{AH - CF}{D} \equiv \dfrac{1}{D}\begin{vmatrix} A & C \\ F & H \end{vmatrix}$, assuming that $D \neq 0$.
1st given eqn in this thm.		Proof: Step #1: $Ax + By = C$ is the 1st assumed constraint.
2nd given eqn in this thm.		Step #2: $Fx + Gy = H$ is the 2nd assumed constraint.
Multiply both equal sides of eqn in step #1 by G.		Step #3: $AGx + BGy = CG$.
Multiply both sides of the eqn in step #2 by $-B$.		Step #4: $-BFx - BGy = -BH$.
Add eqns in steps #3 and #4 using algebra.		Step #5: $(AG - BF)\,x + 0 = CG - BH$.
Multiply both sides of the eqn in step #1 by $-F$.		Step #6: $-AFx - BFy = -CF$.
Multiply both sides of the eqn in step #2 by A.		Step #7: $AFx + AGy = AH$.
Add eqns in steps #6 and #7 using algebra.		Step #8: $0 + (AG - BF)\,y = AH - CF$.

Reason or comment	Statement
Define D (called **Cramer's denominator**) as the common expression in both steps #5 and #8; def (**103**) of determinant.	Step #9: $D \equiv A\,G - B\,F \equiv \begin{vmatrix} A & B \\ F & G \end{vmatrix}$.
Rearrange eqns in steps #5 and #9; def (**103**) of determinant.	Step #10: $x = \dfrac{C\,G - B\,H}{D} \equiv \dfrac{1}{D}\begin{vmatrix} C & B \\ H & G \end{vmatrix}$.
Rearrange eqns in steps #8 and #9; def (**103**) of determinant.	Step #11: $y = \dfrac{A\,H - C\,F}{D} \equiv \dfrac{1}{D}\begin{vmatrix} A & C \\ F & H \end{vmatrix}$.
	This Cramer's rule (**104**) will be used below in statement (**410**).

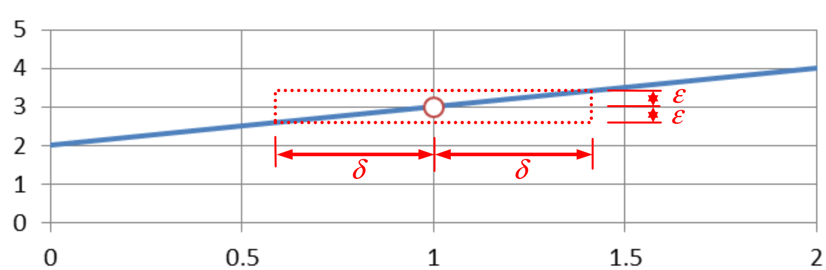
Section 2. Introduce absolute values, limits of functions, and iterative sums

Reason or comment		Statement
Purpose of this Section 2.	**200**)	This Section 2 introduces the **absolute value function** that is probably already familiar to the reader from a study of algebra, the mathematical concept of **limits** that is one of the fundamental ideas on which calculus is based, and the notation for an **iterative summation**. Specifically, limits will be used explicitly in statements (**302**), (**303**), (**304**), and (**306**) in Section 3 below to find slopes of lines that are tangent to a curve with an operation called a "derivative" from differential calculus. For the sake of brevity in this introductory Section 2, only one simple example will be presented of the limit of a particular function at a point for which the function's value is not defined.
Overview of stmts (**201**) through (**207**).		The definition of the absolute value function and proofs of some of its properties follow in statements (**201**) through (**207**) for use in later derivations.
Define the **absolute value** of any real number.	**201**)	(The **absolute value** of q) $\equiv \; \lvert q \rvert \; \equiv \begin{cases} q & \text{if } q \ge 0, \text{ or} \\ -q & \text{if } q < 0 \end{cases}$ for any real q .
Theorem that the absolute value is nonnegative.	**202**)	$\lvert q \rvert \;\; \ge \;\; 0$ for any real q .
Def (**201**); assume $q \ge 0$.		Proof: If $q \ge 0$ then $\lvert q \rvert \; = \; q \;\; \ge \;\; 0$. Otherwise,
Def (**201**); assume $q < 0$.		if $q < 0$ then $\lvert q \rvert \; = -q \; > \; 0$.
Theorem that the square of any real number equals the square of the absolute value of that number.	**203**)	$q^2 \;\; = \;\; \lvert q \rvert^2$ for any real q .
Def (**201**). Square both sides of eqn $q \; = \; \lvert q \rvert$.		Proof: If $q \ge 0$ then $q \; = \; \lvert q \rvert$ so $q^2 \; = \; \lvert q \rvert^2$. Otherwise,
Def (**201**). $q^2 \; = \; (-1)^2 \, q^2$ $= \; \big((-1) \, q\big)^2 \; \equiv \; (-q)^2$; $-q \; = \; \lvert q \rvert$ by def (**201**).		if $q < 0$ then $-q \; = \; \lvert q \rvert$ so $q^2 \; = \; (-q)^2 \; = \; \lvert q \rvert^2$.
Theorem that any real value is less than or equal to its absolute value.	**204**)	$q \;\; \le \;\; \lvert q \rvert$ for any real q .
Def (**201**).		Proof: If $q \ge 0$ then $q \; = \; \lvert q \rvert$. Otherwise,
Assumption; algebra with $q < 0$; def (**201**).		if $q < 0$ then $q \; < \; 0 \; < \; -q \; = \; \lvert q \rvert$, so $q \; < \; \lvert q \rvert$.

Reason or comment	Statement
Theorem that **the absolute value of a product equals the product of the absolute values**.	**205**) $\lvert a\,b\rvert \;=\; \lvert a\rvert\,\lvert b\rvert$ for any real values a and b.
$a\,b \ge 0$ with def (**201**); $a = \lvert a\rvert$ and $b = \lvert b\rvert$ by def (**201**).	Proof: If $a \ge 0$ and $b \ge 0$ then $\lvert a\,b\rvert = a\,b = \lvert a\rvert\,\lvert b\rvert$.
$a\,b < 0$ with def (**201**); algebra; $a = \lvert a\rvert$ and $-b = \lvert b\rvert$ by def (**201**).	If $a \ge 0$ and $b < 0$ then $\lvert a\,b\rvert = -(a\,b) = a\,(-b) = \lvert a\rvert\,\lvert b\rvert$.
$a\,b < 0$ with def (**201**); algebra; $-a = \lvert a\rvert$ and $b = \lvert b\rvert$ by def (**201**).	If $a < 0$ and $b \ge 0$ then $\lvert a\,b\rvert = -(a\,b) = (-a)\,b = \lvert a\rvert\,\lvert b\rvert$.
$a\,b > 0$ with def (**201**); algebra; $-a = \lvert a\rvert$ and $-b = \lvert b\rvert$ by def (**201**).	If $a < 0$ and $b < 0$ then $\lvert a\,b\rvert = a\,b = (-a)\,(-b) = \lvert a\rvert\,\lvert b\rvert$.
Theorem.	**206**) If $r^2 \le s^2$ then $r \le s$ for any nonnegative real values $r \ge 0$ and $s \ge 0$.
	Proof: If $r \ge 0$, $s \ge 0$, and $r^2 \le s^2$ then
Subtract r^2 from both sides of the inequality $r^2 \le s^2$; algebra using the distributive law twice.	$0 \;\le\; s^2 - r^2 \;=\; (s-r)\,\overbrace{(s+r)}^{\ge 0}$, so
$s + r \ge 0$ must be multiplied by $s - r \ge 0$ to make the product nonnegative as required.	$0 \;\le\; s - r$, so
Add r to both sides of the inequality $0 \le s - r$.	$r \;\le\; s$.
This famous theorem is called the **triangle inequality**.	**207**) $\lvert a+b\rvert \;\le\; \lvert a\rvert + \lvert b\rvert$ for any real values a and b. This **triangle inequality** means the absolute value of a sum of 2 real numbers is less than or equal to the sum of the absolute values of those numbers.
Multiply both sides of thm (**204**) by 2 with $q = a\,b$.	Proof: Step #1: $2\,a\,b \;\le\; 2\,\lvert a b\rvert$.
Thm (**203**); algebra.	Step #2: $\lvert a+b\rvert^2 \;=\; (a+b)^2 \;=\; a^2 + 2\,a\,b + b^2$

Reason or comment	Statement
Add $a^2 + b^2$ to both sides of the inequality in step #1.	$\leq\ a^2 + 2\,\lvert ab\rvert + b^2$
Thms (**203**) and (**205**).	$=\ \lvert a\rvert^2 + 2\,\lvert a\rvert\lvert b\rvert + \lvert b\rvert^2$
Algebra.	$=\ (\lvert a\rvert + \lvert b\rvert)^2\,.$
Apply thm (**206**) to the 1st and 6th (last) sides of the inequality in step #2, using thm (**202**).	Step #3: $\lvert a+b\rvert \ \leq\ \lvert a\rvert + \lvert b\rvert\,.$
Define function $f(x)$ for any real value of the argument x except $x = 1$, as explained in stmt (**209**).	**208**) As an example to introduce the concept of limits, consider the function $f(x) \ \equiv\ \dfrac{x^2+x-2}{x-1}\ .$
Show that $f(1)$ is undefined.	**209**) That function is defined for all real values of the argument x where $x \neq 1$, because attempting to evaluate $f(x)$ at $x = 1$ would yield
Eqn (**208**); substitute $x = 1$; arithmetic.	$f(1) \ \equiv\ \left[\dfrac{x^2+x-2}{x-1}\right]_{\text{at } x=1} \ =\ \dfrac{1^2+1-2}{1-1} \ =\ \dfrac{0}{0}\ ,$
Algebra; as examples, $\dfrac{1}{0.1} = 10$, $\dfrac{1}{0.01} = 100$, $\dfrac{1}{0.001} = 1000$, and $1/(-0.001) \ =\ -1000$.	which is not defined because $\dfrac{0}{a} = 0$ for any $a \neq 0$, but $\dfrac{b}{a}$ approaches $\pm$ infinity as a approaches zero (denoted by $a \to 0$) while b is any nonzero constant (that is, any $b \neq 0$), so the **indeterminate value** $\dfrac{0}{0}$ could possibly indicate any value from $-$ infinity through zero to $+$ infinity.
Stmt (**213**) explains ε & δ.	**210**) Here are a graph of $f(x)$ for $0 \leq x \leq 2$ and a table of exact values near $x = 1$.

A graph of $f(x) \equiv \dfrac{x^2+x-2}{x-1}$ as a function of x (undefined at $x = 1$)



$x \ =\ 1+t$	$t \ \equiv\ x-1$	$f(x) \ =\ f(1+t)$
2	1	4
1.1	0.1	3.1
1.01	0.01	3.01
1.001	0.001	3.001
1.0001	0.0001	3.0001
1	**0**	**3 (?)**
0.9999	−0.0001	2.9999
0.999	−0.001	2.999
0.99	−0.01	2.99

Reason or comment		Statement
Define t as the difference between x and 1. Explain that $f(x)$ approaches 3 in the limit as x approaches 1 (which means t approaches zero).	**211**)	In the limit as x **approaches** 1 (**denoted by** $x \to 1$), so that $t \equiv x - 1$ approaches zero (denoted by $\lim_{x \to 1} (x-1) = 0$, **which is read as** "**the limit as** x **approaches** 1 **of** $x - 1$ **equals zero**"), it appears in the graph and table of statement (**210**) just above that the value of function $f(x) = f(1+t)$ approaches 3. Using standard mathematical symbols, this can be written as $\lim_{x \to 1} f(x) = \lim_{t \to 0} f(1+t) = 3$.
Derive a reason why to expect that $f(1)$ approaches 3.	**212**)	To see why this expected conclusion that $\lim_{x \to 1} f(x) = 3$ is actually true, use algebra to simplify function $f(x)$ as
Def of $f(x)$ in stmt (**211**); algebra; algebra.		$$[\,f(x)\,]_{\text{if } x \neq 1} \equiv \frac{x^2 + x - 2}{x - 1} = \frac{x^2 + 2x - x - 2}{x - 1} = \frac{x\,(x+2) - (x+2)}{x - 1}$$
Algebra using the distributive law; algebra that assumes $x \neq 1$.		$$= \frac{(x-1)\ (x+2)}{x - 1} = x + 2\,,$$
The previous eqn, where "$\to$" is read as "approaches"; substitute $x = 1$; arithmetic.		so it is reasonable that $f(1) \to [\,x + 2\,]_{\text{at } x=1} = 1 + 2 = 3$.
This is the standard formal **"epsilon-delta" definition of when the limit of a function** $g(x)$ **equals a real number** p **as the real argument** x **approaches** c.	**213**)	$\lim_{x \to c} g(x) = p$ for some given function $g(x)$ if and only if for any positive real $\varepsilon > 0$ (no matter how small, but not zero) there exists a positive real $\delta > 0$ (typically tiny, but never zero) such that $\lvert\, g(x) - p \,\rvert < \varepsilon$ for all real x such that $0 < \lvert\, x - c \,\rvert < \delta$, where c and p have any real values, x is a real variable, and $g(x)$ must be defined for all those values of $x \neq c$. It is traditional to use the lowercase Greek letters "epsilon" (ε) and "delta" (δ) in this so-called **"epsilon-delta" definition of a limit**.
Restate that **epsilon-delta definition of a limit** using words instead of mathematical symbols.		In other words, for any given positive value denoted by ε, a corresponding positive value denoted by δ can be found such that the absolute value of the difference between function $g(x)$ and the value p of the limit will always be less than any given arbitrarily small allowed error ε provided that the absolute value of the difference between argument x and its target c is less than the maximum allowed deviation δ, assuming $x \neq c$. That is, $g(x)$ can always be forced arbitrarily close to p by keeping x close enough to c. The figure in statement (**210**) above illustrates possible values of ε and δ with argument x approaching its target value $c = 1$, which forces $f(x)$ to approach the value $p = 3$ of its limit.

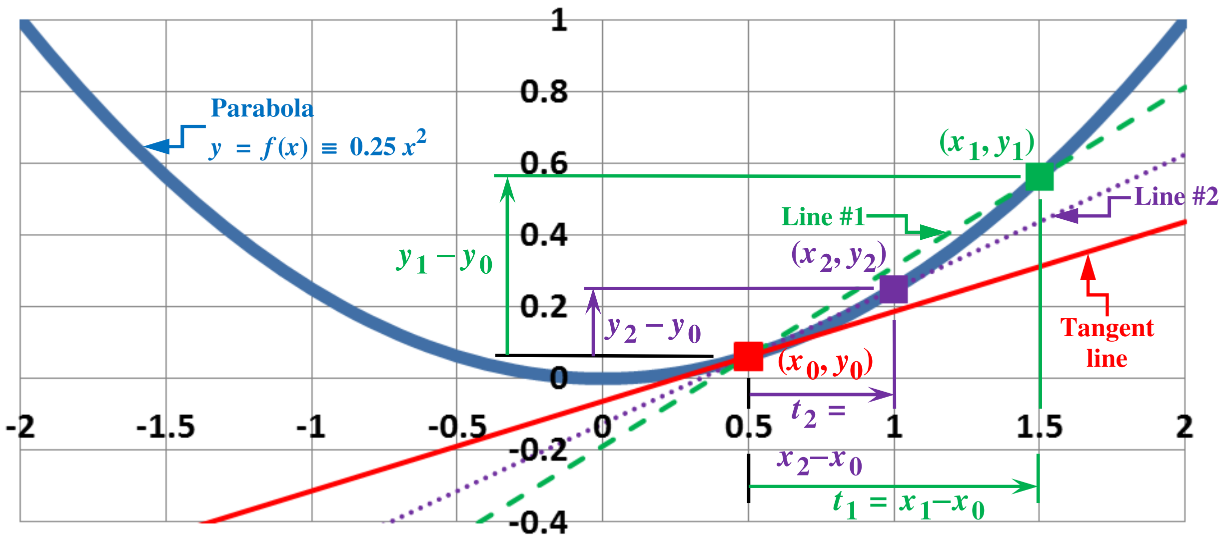
Reason or comment		Statement
Stmts (**214**) through (**216**) introduce sigma $(\sum)$ notation for iterative sum.		Statements (**214**) and (**215**) just below define and present an example of the sigma $(\sum)$ notation for an iterative summation, and theorem (**216**) uses that notation to express an important property of limits.
Define **sigma** $(\sum)$ **notation for an iterative summation** with any positive integer $(n \geq 1)$ number of terms.	**214**)	Define the standard uppercase Greek letter **sigma** $(\sum)$ **notation for the iterative sum** of any n arbitrary functions $h_1(x),\ h_2(x),\ h_3(x),\ \ldots,\ h_n(x)$ by $$\sum_{k=1}^{n} h_k(x) \;\equiv\; h_1(x) + h_2(x) + h_3(x) + \ldots + h_n(x) \qquad \text{for any integer } n \geq 1.$$
An example of def (**214**).	**215**)	As a simple example of an iterative sum using sigma notation, consider
This is read as "the sum of $k\,x$ from k equals 1 to 4" by def (**214**); algebra using the distributive law; arithmetic.		$$\sum_{k=1}^{4} (k\,x) \;\equiv\; 1\,x + 2\,x + 3\,x + 4\,x \;=\; (1+2+3+4)\;x \;=\; 10\,x$$
Algebra; def (**214**).		$$=\; x\;(1+2+3+4) \;\equiv\; x\sum_{k=1}^{4} k\;.$$
Theorem that **the limit of the sum of n arbitrary functions equals the sum of the limits of those functions**.	**216**)	$$\lim_{x\to c}\left(\sum_{k=1}^{n} h_k(x)\right) \;\equiv\; \lim_{x\to c}\big(h_1(x) + h_2(x) + \ldots + h_n(x)\big)$$ $$=\; \left(\lim_{x\to c} h_1(x)\right) + \left(\lim_{x\to c} h_2(x)\right) + \ldots + \left(\lim_{x\to c} h_n(x)\right)$$
See def (**214**) of sigma $(\sum)$ notation for iterative summations.		$$\equiv\; \sum_{k=1}^{n}\left(\lim_{x\to c} h_k(x)\right)$$ for any $n \geq 1$ arbitrary functions $h_1(x),\ h_2(x),\ h_3(x),\ \ldots,\ h_n(x)$ whose limits exist as argument x approaches the given value c.
Define $p_1,\ p_2,\ \ldots,\ p_n$.		Proof: Let $p_1,\ p_2,\ \ldots,\ p_n$ respectively denote the limits of the functions $h_1(x),\ h_2(x),\ \ldots,\ h_n(x)$ as argument x approaches the given real value c; that is, define the notation $$p_k \;\equiv\; \lim_{x\to c} h_k(x) \qquad \text{for each integer } k \in \{\,1, 2, \ldots, n\,\}.$$
Explain what must be derived to complete this proof.		Then to prove this theorem it must be shown that $$\lim_{x\to c}\left(\sum_{k=1}^{n} h_k(x)\right) \;=\; p_1 + p_2 + \ldots + p_n \;\equiv\; \sum_{k=1}^{n} p_k\;.$$

Reason or comment	**Statement**
Select any $\varepsilon > 0$.	Choose any arbitrarily small but nonzero value of $\varepsilon > 0$. By definition (**213**) of a limit, a corresponding value of δ must be found such that
Describe the condition that δ must satisfy for the chosen value of ε.	$\left\| \left(\sum_{k=1}^{n} h_k(x) \right) - \left(\sum_{k=1}^{n} p_k \right) \right\| < \varepsilon$ for all $x \neq c$ such that $\| x - c \| < \delta$.
Def (**213**) of a limit.	The definition $p_k \equiv \lim_{x \to c} h_k(x)$ for each $k \in \{ 1, 2, \ldots, n \}$ means that values of δ_1, δ_2, and δ_n can be found such that $\| h_k(x) - p_k \| < \dfrac{\varepsilon}{n}$ for all values of $x \neq c$ where $\| x - c \| < \delta_k$.
Choose a value of δ.	Let δ equal the minimum of the values δ_1, δ_2, and δ_n.
$\delta \equiv \min(\delta_1, \delta_2, \ldots, \delta_n)$.	For any $k \in \{ 1, 2, \ldots, n \}$, it will be true that $\delta \le \delta_k$, and so
$\| x - c \| < \delta \le \delta_k$.	if $\| x - c \| < \delta$ and $x \neq c$ then $\| h_k(x) - p_k \| < \dfrac{\varepsilon}{n}$. Therefore,
Rearrange the order of terms by repeatedly applying the commutative and associative laws of algebra.	$\left\| \left(\sum_{k=1}^{n} h_k(x) \right) - \left(\sum_{k=1}^{n} p_k \right) \right\| = \left\| \sum_{k=1}^{n} \left(h_k(x) - p_k \right) \right\|$
Apply the triangle inequality (**207**) n times.	$\le \sum_{k=1}^{n} \left\| h_k(x) - p_k \right\|$
Apply $0 \le \| h_k(x) - p_k \| < \dfrac{\varepsilon}{n}$ with algebra; def (**214**) of an iterative summation.	$< \sum_{k=1}^{n} \dfrac{\varepsilon}{n} \equiv \overbrace{\dfrac{\varepsilon}{n} + \dfrac{\varepsilon}{n} + \cdots + \dfrac{\varepsilon}{n}}^{\text{sum of } n \text{ identical terms}}$
The definition of multiplication by n; algebra.	$\equiv \dfrac{\varepsilon}{n}(n) = \varepsilon$.

Section 3. Slope of a line tangent to a parabola at any point

Reason or comment	Statement
Overview of Section 3. Although several of the important techniques and notations introduced in this section are based on differential calculus, it is intended that a careful reader who is not already familiar with calculus should be able to understand these concepts from explanations here.	**300**) In this Section 3, statement (**301**) defines the slope and the y-axis intercept of a straight line, statement (**302**) defines the slope of a line that is tangent at some point to the curve drawn by plotting a function and describes that slope as a derivative from differential calculus, statement (**303**) uses that definition to derive an expression for the slope of a line that is tangent to the simple parabolic curve $f(x) \equiv A\,x^2$ at any x, statement (**304**) derives an expression for the slope of the line that is tangent to the alternative parabolic curve for function $g(x) \equiv (B\,x + C)^2$ at any x, statement (**305**) describes how solving for the argument's value required to make a derivative equal zero can be used to find local minimum and maximum points of a function, and theorem (**306**) states that the slope of the sum of any n functions equals the sum of the slopes of those functions at any given value of their argument.
Define **slope and y-axis intercept of a (straight) line**. "Δ" is the upper-case Greek letter "delta".	**301**) The **slope of a straight line** (such as a plot of the function $y = m\,x + b$) equals the ratio of the change of the y coordinate (which is called the **rise** of part of the line and is traditionally denoted by Δy) divided by the change of the x coordinate (which is called the **run** of the part and is denoted by Δx) .
The coordinates of all points (x, y) on the line must have $y = m\,x + b$.	Consider the part between any two points (x_1, y_1) and (x_2, y_2) on that line, so $y_1 = m\,x_1 + b$ and $y_2 = m\,x_2 + b$.
Show that the slope of the line is m .	Then the **slope of that line equals** m for any values of x_1 and x_2 (assuming the slope is finite so the line is not exactly vertical, and that $x_1 \neq x_2$) because
Def of slope; defs of rise and run; defs of Δx and Δy ; $y = m\,x + b$ for all points on the line.	$$\text{slope} \equiv \frac{\text{rise}}{\text{run}} \equiv \frac{\Delta y}{\Delta x} \equiv \frac{y_2 - y_1}{x_2 - x_1} = \frac{(m\,x_2 + b) - (m\,x_1 + b)}{x_2 - x_1}$$
Algebra; algebra.	$$= \frac{m\,(x_2 - x_1)}{x_2 - x_1} = m\,.$$
Show that the y-axis intercept of the line is b .	The ***y*-axis intercept of that line equals** b because the y coordinate at which the line crosses the y-axis (that is, where $x = 0$, again assuming the line is not exactly vertical so the slope of the line is finite) is
Def of the y-axis intercept.	(y-axis intercept) $\equiv$ (y coordinate of the point on the line where $x = 0$)
$y = m\,x + b$ for all points (x, y) on the line; substitute $x = 0$; algebra.	$$\equiv [\,m\,x + b\,]_{\text{at } x=0} = m\,0 + b = b\,.$$

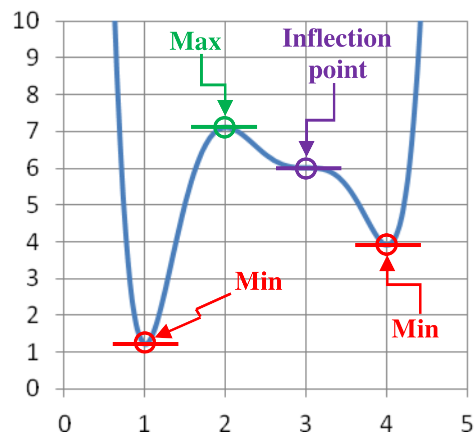
Reason or comment	**Statement**
Define the **slope of a line that is tangent** at some point $(x, f(x))$ **to** the curve from plotting the **function** $f(x)$. That slope is called the **derivative** of function $f(x)$ in the terminology of **differential calculus**.	**302**) Consider any point $(x, f(x))$ on the curve produced by plotting a given function $f(x)$ at some given value(s) of x. The slope of the line tangent to that curve with such a value of x can be approximated as the slope of the line between two points $(x, f(x))$ and $(x+t,\ f(x+t))$ on the curve for a small value of $t \equiv \Delta x$. In fact, the **slope of the line tangent to that curve** at the point $(x, f(x))$ is defined as the slope of the line described in the previous sentence in the limit as $t \equiv \Delta x$ approaches zero. That means
	(slope of line tangent to the curve of function $f(x)$ at a given point $(x, f(x))$)
See defs (**301**) and (**213**); $\Delta f(x) \equiv f(x+t) - f(x)$ corresponds to $\Delta x \equiv (x+t) - t$; algebra.	$\equiv \lim\limits_{\Delta x \to 0}\left(\dfrac{\Delta f}{\Delta x}\right) \equiv \lim\limits_{t \to 0}\left(\dfrac{f(x+t) - f(x)}{(x+t) - x}\right) = \lim\limits_{t \to 0}\left(\dfrac{f(x+t) - f(x)}{t}\right)$
Define a simple **notation for a derivative**; define a more explicit notation for a derivative.	$\equiv \dfrac{\mathrm{d}f}{\mathrm{d}x} \equiv \dfrac{\mathrm{d}}{\mathrm{d}x} f(x)$
This is how to read those two expressions in words.	$\equiv$ (the **derivative** of $f(x)$ with respect to x, evaluated at given argument x).
Describe two notations in the previous equation for a derivative of function $f(x)$ with respect to its argument x. The notation $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ represents a limit rather than simply a ratio calculated by dividing two tiny numbers.	That definition of a so-called "**derivative**" is a key concept of **differential calculus**. In the 5th equal side of the previous equation, the derivative is represented by the notation $\dfrac{\mathrm{d}f}{\mathrm{d}x}$, which is mnemonic for the ratio of an infinitesimally tiny (that is, a "**differential** $\mathrm{d}f$") change in the value of function $f(x)$ divided by the corresponding infinitesimally small (or "differential $\mathrm{d}x$") change in the value of argument x. However, that notation $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ and the equivalent notation $\dfrac{\mathrm{d}}{\mathrm{d}x} f(x)$ actually represent the limit $\lim\limits_{t \to 0}\left(\dfrac{f(x+t) - f(x)}{t}\right)$ rather than a simple ratio of two tiny numbers.
Consider function $f(x)$ for a **"vertical" parabola** that is symmetric about the y axis with its minimum (if $A > 0$) or maximum (if $A < 0$) at the origin $(0, 0)$.	**303**) As an example, consider the particular function $f(x) \equiv A\,x^2$ for all real argument values x, whose plot is called a **parabola** (which is one of the **conic sections**, together with a **circle**, an **ellipse**, and a **hyperbola**). A plot of that function $f(x)$ with the specific parameter value $A = 0.25$ is the thick blue solid parabola shown in the following figure.



Reason or comment	Statement
	The goal here is to determine the slope of the thin red solid line that is tangent to the parabola at the red point $(x_0, y_0) = (0.5, 0.0625)$.
The slope of line #1 crudely approximates the slope of the tangent.	As a first crude approximation, consider the green dashed line #1 passing through the red tangent point $(x_0, y_0) = (0.5, 0.0625)$ and the green point $(x_1, y_1) = (1.5, 0.5625)$, both of which are on the parabola. That line #1 has a slope of
Def (**301**); $t_1 \equiv x_1 - x_0$.	(slope of green dashed line #1) $\equiv \dfrac{y_1 - y_0}{x_1 - x_0} \equiv \dfrac{f(x_0 + t_1) - f(x_0)}{t_1}$
Substitute values; arithmetic; arithmetic.	$= \dfrac{0.5625 - 0.0625}{1.5 - 0.5} = \dfrac{0.5}{1} = 0.5$.
The slope of line #2 better approximates the slope of the tangent.	As a second, somewhat more accurate approximation, consider the purple dotted line #2 passing through the red tangent point $(x_0, y_0) = (0.5, 0.0625)$ and the closer purple point $(x_2, y_2) = (1, 0.25)$, which has a slope of
Def (**301**); $t_2 \equiv x_2 - x_0$.	(slope of purple dotted line #2) $\equiv \dfrac{y_2 - y_0}{x_2 - x_0} \equiv \dfrac{f(x_0 + t_2) - f(x_0)}{t_2}$
Substitute values; arithmetic; arithmetic.	$= \dfrac{0.25 - 0.0625}{1 - 0.5} = \dfrac{0.1875}{0.5} = 0.375$.

Reason or comment	Statement
Find the exact slope of the tangent line using a limit as $t \to 0$.	The exact value of the slope of the thin red solid tangent line can be obtained similarly by calculating the slope of the line through the point $(x_0, y_0) = (0.5, 0.0625)$ and another point $(x_0+t,\ f(x_0+t))$ on the parabola in the limit as t approaches zero. That limit yields
	(exact slope of the thin red solid tangent line at the point $(x_0,\ f(x_0))$)
Def (**302**) of a derivative; def (**302**).	$\equiv \left[\dfrac{df}{dx}\right]_{\text{at } x=x_0} \equiv \lim\limits_{t\to 0}\left(\dfrac{f(x_0+t)-f(x_0)}{t}\right)$
$f(x) \equiv A\,x^2$ for this parabola; algebra.	$\equiv \lim\limits_{t\to 0}\left(\dfrac{A\,(x_0+t)^2 - A\,x_0^2}{t}\right) = \lim\limits_{t\to 0}\left(\dfrac{A\,(x_0^2+2\,x_0\,t+t^2) - A\,x_0^2}{t}\right)$
Algebra; algebra.	$= \lim\limits_{t\to 0}\left(\dfrac{2\,A\,x_0\,t + A\,t^2}{t}\right) = \lim\limits_{t\to 0}\left(2\,A\,x_0 + A\,t\right)$
$A\,t \to 0$ as $t \to 0$; substitute values for this example; arithmetic.	$= 2\,A\,x_0 = 2\,(0.25)\,0.5 = 0.25$.
Find the slope of parabola $g(x) \equiv (B\,x+C)^2$.	**304**) Similarly, for any real constants B and C, the exact slope of a line tangent to the parabola $g(x) \equiv (B\,x+C)^2$ at the point $(x,\ g(x))$ for any real x is
Def (**302**); def (**302**).	(slope of parabola $g(x)$) $\equiv \dfrac{dg}{dx} \equiv \lim\limits_{t\to 0}\left(\dfrac{g(x+t)-g(x)}{t}\right)$
See the proof just below.	$= 2\,B\,(B\,x+C)$.
$g(x) \equiv (B\,x+C)^2$ for this parabola.	Proof: $\lim\limits_{t\to 0}\dfrac{g(x+t)-g(x)}{t} \equiv \lim\limits_{t\to 0}\dfrac{(B\,(x+t)+C)^2 - (B\,x+C)^2}{t}$
Algebra.	$= \lim\limits_{t\to 0}\dfrac{\left(B^2\,(x+t)^2 + 2\,B\,(x+t)\,C + C^2\right) - (B^2\,x^2 + 2\,B\,x\,C + C^2)}{t}$
Algebra.	$= \lim\limits_{t\to 0}\dfrac{B^2(x^2+2xt+t^2) + 2BxC + 2BtC + C^2 - B^2x^2 - 2BxC - C^2)}{t}$
Algebra; algebra.	$= \lim\limits_{t\to 0}\dfrac{2B^2xt + B^2t^2 + 2BtC}{t} = \lim\limits_{t\to 0}(2B^2x + B^2t + 2BC)$
$B^2\,t \to 0$ as $t \to 0$; algebra with the distributive law.	$= 2\,B^2\,x + 2\,B\,C = 2\,B\,(B\,x+C)$.

Reason or comment		Statement
Explain how a derivative can be used to **find local minimum or maximum points of a function**.	**305**)	One important application of derivatives (that is, slopes of lines tangent to a curve) is using them to find the local minimum and local maximum points of a function. As shown in the following example graph, **derivatives have the value zero** (that is, **tangent lines are horizontal** because they have zero slope) at each point **where a curve reaches a local minimum** (where the curve is "concave up"), **a local maximum** (where the curve is "concave down"), **or an inflection point** (where the curve changes from concave up to concave down as shown, or else from concave down to concave up).



This curve is a plot of the function
$\boldsymbol{h(x) = x^6 - 15.6\,x^5 + 97.5\,x^4 - 310\,x^3 + 522\,x^2 - 432\,x + 138.3}$ **.**

It has two local minima at the points (1, 1.2) and (4, 3.9) .

It has one local maximum at the point (2, 7.1) .

It has one inflection point at (3, 6) .

The slopes of the tangent lines (that is, evaluating the derivative of the curve) at all of those four points equal zero.

Although it is not proved here, the derivative of $h(x)$ is

$$\frac{\mathrm{d}}{\mathrm{d}x} h(x) \;=\; 6\,x^5 - 78\,x^4 + 390\,x^3 - 930\,x^2 + 1044\,x - 432$$

$$=\; 6\,(x-1)\,(x-2)\,(x-3)^2\,(x-4)\,,$$

which obviously equals zero for each $x \in \{\,1, 2, 3, 4\,\}$.

Reason or comment	Statement
Find the minimum of the parabolic function $f(x) = A\,x^2 = 0.25\,x^2$ from stmt (**303**).	As a second example of this technique, statement (**303**) showed a graph of the parabolic function $f(x) = A\,x^2 = 0.25\,x^2$ and proved that its derivative is $\frac{\mathrm{d}}{\mathrm{d}x} f(x) = 2\,A\,x = 0.5\,x$. That derivative equals zero when $x = 0$, which is at the minimum value of the curve plotted in statement (**303**).
Find the minimum of the parabolic function $g(x) = (B\,x + C)^2$ from stmt (**304**). $[2\,B\,(B\,x + C)^2]_{\text{at }x=-C/B} = 0.$ $g(-C/B) \equiv [(B\,x + C)^2]_{\text{at }x=-C/B} = 0.$	As a third example of this technique, statement (**304**) proved that the derivative of the function $g(x) = (B\,x + C)^2$ is $\frac{\mathrm{d}}{\mathrm{d}x} g(x) = 2\,B\,(B\,x + C)$. That derivative equals zero at $x = -\frac{C}{B}$, where value of the original function $g(x)$ is $g\left(-\frac{C}{B}\right) = 0$, which is at the minimum value of the parabolic curve corresponding to the function $g(x) \ge 0$ for all real x .

Reason or comment	Statement
Theorem that **the slope of the sum of any** n **functions equals the sum of the slopes** of those functions at any given value of their argument, using def (**302**). See def (**214**) of sigma $(\sum)$ notation for sums, and see def (**302**) of a derivative.	**306**) $\dfrac{d}{dx}\sum_{k=1}^{n} h_k(x) \equiv \dfrac{d}{dx}\big(h_1(x) + h_2(x) + \ldots + h_n(x)\big)$ $= \dfrac{d}{dx}h_1(x) + \dfrac{d}{dx}h_2(x) + \ldots + \dfrac{d}{dx}h_n(x) \equiv \sum_{k=1}^{n}\dfrac{d}{dx}h_k(x)$ for any $n \ge 1$ functions $h_1(x),\ h_2(x),\ \ldots,\ h_n(x)$, assuming that all of those derivatives exist. This theorem means that **the derivative of a sum of any** n **functions equals the sum of the derivatives** of those n functions.
Define function $h(x)$ for use in this proof.	Proof: Define the function $h(x) \equiv \sum_{k=1}^{n} h_k(x)$. The derivative of that function is
Substitute $h(x) \equiv \sum_{k=1}^{n} h_k(x)$; def (**302**) of a derivative.	$\dfrac{d}{dx}\sum_{k=1}^{n} h_k(x) \equiv \dfrac{d}{dx}h(x) \equiv \lim_{t\to 0}\dfrac{h(x+t)-h(x)}{t}$
Substitute $h(x) \equiv \sum_{k=1}^{n} h_k(x)$.	$\equiv \lim_{t\to 0}\dfrac{\left(\sum_{k=1}^{n} h_k(x+t)\right) - \left(\sum_{k=1}^{n} h_k(x)\right)}{t}$
Rearrange the order of terms in the sums using the associative and commutative laws of algebra.	$= \lim_{t\to 0}\dfrac{\sum_{k=1}^{n}\big(h_k(x+t)-h_k(x)\big)}{t}$
The distributive law of algebra.	$= \lim_{t\to 0}\left(\sum_{k=1}^{n}\dfrac{h_k(x+t)-h_k(x)}{t}\right)$
Thm (**216**); def (**302**).	$= \sum_{k=1}^{n}\left(\lim_{t\to 0}\dfrac{h_k(x+t)-h_k(x)}{t}\right) \equiv \sum_{k=1}^{n}\dfrac{d}{dx}h_k(x)$.

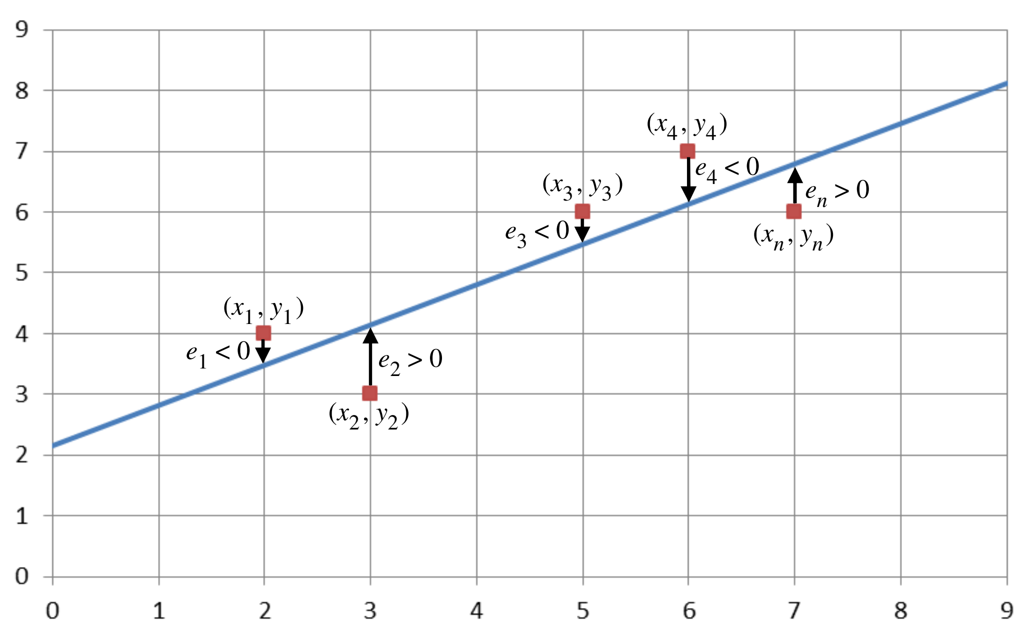
Section 4. Determine a linear trendline

Reason or comment	Statement
Purpose of this Section 4. Stmt (**404**) explains a quantitative interpretation of the "best" line.	**400**) Using the mathematical principles presented in Section 1 though Section 3 above, this Section 4 derives the equation of the "best" straight line to approximate a set of any $n \geq 2$ pairs of (x, y) coordinates for known sample points. Microsoft Excel uses this method to calculate a linear trendline.
Describe the given sample points and the straight trendline.	**401**) Assume coordinates of $n \geq 2$ sample points (x_1, y_1), (x_2, y_2), …, (x_n, y_n) are known. The goal is to determine the slope m and the y-axis intercept b for the "best" straight line with the function $y = f(x) \equiv m\,x + b$, which should pass as near as possible to the n given sample points.
Define e_k to be the signed error from k^{th} sample point to the trendline.	**402**) For each sample point number $k \in \{1, 2, 3, \ldots, n\}$, define the algebraically signed error from the k^{th} sample point (x_k, y_k) to the point $(x_k,\ m\,x_k + b)$ on the trendline (which will eventually be calculated) to be $e_k \equiv m\,x_k + b - y_k \equiv$ (vertical error from k^{th} sample point to trendline).
Interpret the algebraic sign of error e_k.	Therefore, error e_k is positive if the trendline is plotted above the k^{th} sample point, e_k is zero if the trendline passes through that sample point, or else e_k is negative if the trendline is plotted below that sample point.
	403) Consider the following example of $n = 5$ sample points and a linear trendline.

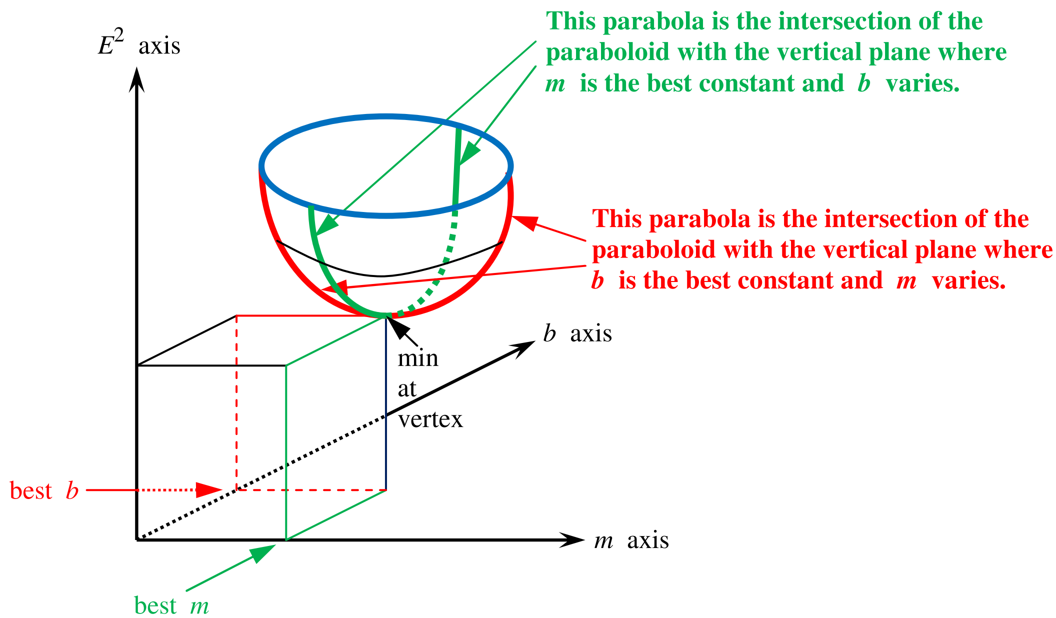
The coordinates of the $n = 5$ data samples follow:

k	x_k	y_k
1	2	4
2	3	3
3	5	6
4	6	7
5	7	6

Statement (**411**) calculates the linear trendline.



Reason or comment	Statement
Why would minimizing the sum of all n signed errors not yield a good trendline? See def (**214**) of sigma (Σ) notation for iterative summations.	**404**) The "best" linear trendline can be found by minimizing some measure of the overall error between the given sample points and the trendline. The summation $\sum_{k=1}^{n} e_k$ of all n signed error terms defined just above would not be an appropriate measure of total error because that sum becomes more and more negative when a candidate trendline passes farther and farther below the sample points toward negative infinity. Therefore, an attempt to minimize that simple sum of errors would drive the trendline toward $-$ infinity regardless of what sample points are given.
Minimizing the sum of absolute values of all n signed errors would yield a good trendline.	An appropriate measure of total error to be minimized must increase whenever a candidate trendline gets farther away from given sample points in some reasonable sense. Such a measure of total error might be the sum of the absolute values of the errors in definition (**402**).
Actually the **method of least squares** will be used to find the "best" linear trendline by minimizing this sum of the squares of all n signed errors.	However, it is computationally easier and is generally more effective for reducing especially errors with relatively larger absolute values to determine the "best" linear trendline by minimizing the "total squared error" defined as the sum of the squares of the signed errors by $$E^2 \quad \equiv \quad \sum_{k=1}^{n} e_k^2 \quad \equiv \quad \sum_{k=1}^{n} (m\,x_k + b - y_k)^2 \quad \equiv \quad \text{(total squared error).}$$
Minimizing the sum of e_k^2 tends to reduce large absolute errors much more than minimizing the sum of $\lvert e_k \rvert$ would.	For example, an absolute error of $\lvert e_k \rvert = 10$ increases this total squared error by $10^2 = 100$ times as much as an absolute error of $\lvert e_k \rvert = 1$ does, whereas it would increase the sum of $\lvert e_k \rvert$ values by only 10 times as much. That indicates why the total squared error E^2 emphasizes larger absolute errors much more than the sum of absolute errors does.
See definitions of the slope m and the y-axis intercept b in stmt (**301**). See def (**404**) of the total squared error E^2. The figure in this statement will be used to explain how the method of least squares is derived in stmts (**406**) through (**410**) just below.	**405**) Therefore, the **method of least squares** will be used to determine the slope m and the y-axis intercept b of the "best" trendline $f(x) \equiv m\,x + b$ by minimizing the total squared error E^2. To understand how to choose values of m and b to minimize E^2, start by considering how E^2 changes as a function of m and b. Based on the equation at the end of statement (**404**) just above and comparing it with statement (**304**) in Section 3, a plot of E^2 while varying b but holding m constant is a parabolic curve, and also a plot of E^2 while varying m but holding b constant is another parabolic curve. That is shown in the following sketch of E^2 as a function of both m and b. The bowl-like surface is called a "**paraboloid**"; or more precisely, this is an "elliptic paraboloid" to distinguish it from a "hyperbolic paraboloid". (See webpage http://mathworld.wolfram.com/Paraboloid.html.) The minimum value of E^2 is at the bottom point (which is called the "**vertex**") of the paraboloid.



Reason or comment	Statement
Determine a first constraint on m and b that must be satisfied at the minimum point of the paraboloid in figure (**405**) when m is held constant while b varies.	**406**) As described in statement (**305**), the slope equals zero for a line that is tangent to a curve at a point where that curve has a local minimum. That means the derivative of the function represented by that curve equals zero at the local minimum point. Therefore, the minimum point on the parabola which is the intersection of the paraboloid surface of E^2 with a vertical plane parallel to the b and E^2 axes where m is held constant (possibly at its best value) while b varies must satisfy the constraint
See stmt (**305**); substitute $E^2 \equiv \sum_{k=1}^{n} (m\,x_k + b - y_k)^2$.	$0 = \left[\frac{\mathrm{d}}{\mathrm{d}b} E^2\right]_{\substack{\text{vary } b \text{ with} \\ m\,=\,\text{constant}}} \equiv \frac{\mathrm{d}}{\mathrm{d}b} \sum_{k=1}^{n} (m\,x_k + b - y_k)^2$
Thm (**306**); thm (**304**).	$= \sum_{k=1}^{n} \frac{\mathrm{d}}{\mathrm{d}b} (m\,x_k + b - y_k)^2 = \sum_{k=1}^{n} 2\,(m\,x_k + b - y_k)$
Algebra using the distributive law.	$= 2 \sum_{k=1}^{n} (m\,x_k + b - y_k)$
Rearrange the order of terms by applying the associative, commutative, and distributive laws of algebra.	$= 2 \left(m \sum_{k=1}^{n} x_k + \sum_{k=1}^{n} b - \sum_{k=1}^{n} y_k \right)$

Reason or comment	Statement
Definition of multiplication by n.	$= \; 2 \left(m \sum_{k=1}^{n} x_k \; + \; b\,n \; - \; \sum_{k=1}^{n} y_k \right) .$
Determine a second constraint on m and b that must be satisfied at the minimum point of the paraboloid in figure (**405**) when b is held constant while m varies.	**407**) Similarly as described in statement (**305**), the minimum point on the parabola which is the intersection of the paraboloid surface of E^2 with a vertical plane parallel to the m and E^2 axes where b is held constant (possibly at its best value) while m varies must satisfy the constraint
See stmt (**305**); substitute $E^2 \equiv \sum_{k=1}^{n} (m\,x_k + b - y_k)^2$.	$0 \; = \; \left[\frac{\mathrm{d}}{\mathrm{d}m} E^2 \right]_{\substack{\text{vary } m \text{ with} \\ b\,=\,\text{constant}}} \; \equiv \; \frac{\mathrm{d}}{\mathrm{d}m} \sum_{k=1}^{n} (m\,x_k + b - y_k)^2$
Thm (**306**); thm (**304**).	$= \; \sum_{k=1}^{n} \frac{\mathrm{d}}{\mathrm{d}m} (m\,x_k + b - y_k)^2 \; = \; \sum_{k=1}^{n} 2\,x_k\,(m\,x_k + b - y_k)$
Algebra using the distributive law.	$= \; 2 \sum_{k=1}^{n} (m\,x_k^2 + b\,x_k - x_k\,y_k)$
Rearrange the order of terms by repeatedly applying the associative, commutative, and distributive laws of algebra.	$= \; 2 \left(m \sum_{k=1}^{n} x_k^2 \; + \; b \sum_{k=1}^{n} x_k \; - \; \sum_{k=1}^{n} x_k\,y_k \right) .$
	The two constraints on m and b in equations (**406**) and (**407**) just above can be rearranged by algebra into equations (**408**) and (**409**) just below.
Apply algebra to the 1st and 8th (last) sides of eqn (**406**).	**408**) $\sum_{k=1}^{n} y_k \; = \; m \sum_{k=1}^{n} x_k \; + \; b\,n \; .$
Apply algebra to the 1st and 7th (last) sides of eqn (**407**).	**409**) $\sum_{k=1}^{n} x_k\,y_k \; = \; m \sum_{k=1}^{n} x_k^2 \; + \; b \sum_{k=1}^{n} x_k \; .$
State the method to **compute** m **and** b **for the best trendline**.	**410**) Cramer's rule in theorem (**104**) can be used to solve the two simultaneous linear equations (**408**) and (**409**) to **compute the best values of the slope** m **and the** ***y*****-axis intercept** b **by the algorithm**

Reason or comment	Statement
Apply Cramer's rule (**104**) to eqns (**408**) and (**409**).	$D \equiv (\text{Cramer's denominator}) = \begin{vmatrix} \sum_{k=1}^{n} x_k & n \\ \sum_{k=1}^{n} x_k^2 & \sum_{k=1}^{n} x_k \end{vmatrix}$
Evaluate the 2×2 determinant by def (**103**).	$\equiv \left(\sum_{k=1}^{n} x_k \right)^2 - n \sum_{k=1}^{n} x_k^2 \;;$
Apply Cramer's rule (**104**) to eqns (**408**) and (**409**).	$\text{best } m = \frac{1}{D} \begin{vmatrix} \sum_{k=1}^{n} y_k & n \\ \sum_{k=1}^{n} x_k\, y_k & \sum_{k=1}^{n} x_k \end{vmatrix}$
Evaluate the 2×2 determinant by def (**103**).	$= \frac{1}{D} \left(\left(\sum_{k=1}^{n} x_k \right) \left(\sum_{k=1}^{n} y_k \right) - n \left(\sum_{k=1}^{n} x_k\, y_k \right) \right) \;;$
Apply Cramer's rule (**104**) to eqns (**408**) and (**409**).	$\text{best } b = \frac{1}{D} \begin{vmatrix} \sum_{k=1}^{n} x_k & \sum_{k=1}^{n} y_k \\ \sum_{k=1}^{n} x_k^2 & \sum_{k=1}^{n} x_k\, y_k \end{vmatrix}$
Evaluate the 2×2 determinant by def (**103**).	$= \frac{1}{D} \left(\left(\sum_{k=1}^{n} x_k \right) \left(\sum_{k=1}^{n} x_k\, y_k \right) - \left(\sum_{k=1}^{n} x_k^2 \right) \left(\sum_{k=1}^{n} y_k \right) \right) .$
Show how to solve for the trendline in figure (**403**).	**411**) The algorithm in statement (**410**) just above can be used to compute the linear trendline from the sample points shown in figure (**403**) as follows.

k	x_k	y_k	x_k^2	$x_k\, y_k$	$m\, x_k + b$	$e_k \equiv m\, x_k + b - y_k$
1	2	4	4	8	3.476744186	−0.523255814
2	3	3	9	9	4.139534884	1.139534884
3	5	6	25	30	5.465116279	−0.534883721
4	6	7	36	42	6.127906977	−0.872093023
5	7	6	49	42	6.790697674	0.790697674

$n =$	$\sum_{k=1}^{n} x_k =$	$\sum_{k=1}^{n} y_k =$	$\sum_{k=1}^{n} x_k^2 =$	$\sum_{k=1}^{n} x_k\, y_k =$
5	23	26	123	131

Reason or comment	Statement
Cramer's denominator is used to calculate m and b.	$D \;=\; \left(\sum_{k=1}^{n} x_k\right)^2 \;-\; n \sum_{k=1}^{n} x_k^2$
Substitute values; arithmetic.	$=\; 23^2 \;-\; 5\,(123) \;=\; -86\,;$
The slope of the best linear trendline.	$m \;=\; \dfrac{1}{D}\left(\left(\sum_{k=1}^{n} x_k\right)\left(\sum_{k=1}^{n} y_k\right) \;-\; n\left(\sum_{k=1}^{n} x_k\, y_k\right)\right)$
Substitute values; arithmetic; arithmetic.	$=\; \dfrac{23\,(26) \;-\; 5\,(131)}{-86} \;=\; \dfrac{57}{86} \;\approx\; 0.662790698\,;$
The y-axis intercept of the best linear trendline.	$b \;=\; \dfrac{1}{D}\left(\left(\sum_{k=1}^{n} x_k\right)\left(\sum_{k=1}^{n} x_k\, y_k\right) \;-\; \left(\sum_{k=1}^{n} x_k^2\right)\left(\sum_{k=1}^{n} y_k\right)\right)$
Substitute values; arithmetic; arithmetic.	$=\; \dfrac{23\,(131) \;-\; 123\,(26)}{-86} \;=\; \dfrac{185}{86} \;\approx\; 2.151162791\,.$
Sample points with at least two distinct x coordinates are needed.	**412**) The least squares algorithm in statement (**410**) cannot calculate any linear trendline if only $n = 1$ sample point is given or if all $n \geq 2$ sample points have the same x coordinate, because then Cramer's denominator is
Stmt (**410**); assume $x_k = x_1$ for all $1 \leq k \leq n$.	$D \;=\; \left(\sum_{k=1}^{n} x_k\right)^2 - n \sum_{k=1}^{n} x_k^2 \;=\; \left(\sum_{k=1}^{n} x_1\right)^2 - n \sum_{k=1}^{n} x_1^2$
Algebra; algebra.	$=\; (n\, x_1)^2 \;-\; n\,(n\, x_1^2) \;=\; 0\,,$
See stmt (**209**).	but division by a denominator equal to zero is undefined.
List mathematical topics introduced in this document.	**413**) In summary, this document introduced the following important mathematical topics in an effort to derive an algorithm to calculate the slope m and the y-axis intercept b of a linear trendline $y = m\,x + b$ that approximates an arbitrary number $n \geq 2$ of discrete samples $(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)$. 1. **Matrix theory** and **Cramer's rule** from **linear algebra**. 2. Properties of the **absolute value function**. 3. The concept of **limits of a function** at values of the argument where the function may or may not be defined. 4. **Sigma** $(\sum)$ **notation for an iterative summation**. 5. The **slope of the line that is tangent to a curve** (such as a **parabola**) at a point on that curve, using a limit. 6. Express the value of that slope as a **derivative** of the function for that curve from **differential calculus**. 7. How to use slopes (that is, derivatives) to find each local **minimum (or maximum or inflection point) of a function**. 8. How to determine the **best linear trendline** by minimizing the **least square error** for the given sample data points.